图书情报工作
第 62 卷 第 4 期 2018 年 2 月

中文科研论文未被引探索Ⅰ:外部特征影响研究*
——以图书馆情报与文献学为例

韩毅 伍玉 申东阳 况书梅 袁庆
西南大学计算机与信息科学学院 重庆 400715

摘要:[**目的/意义**]探索论文未被引现象是引文分布研究中不可或缺的部分,不仅有利于丰富和扩展计量学的研究范畴,也有利于识别文献未被引的产生机制和最大限度避免科研资源浪费与提升科学交流效率。[**方法/过程**]以 CSSCI 为来源数据库,以图书馆情报与文献学为样本学科,随机选择 200 名学者为样本,获取这些学者的第一作者论文及相关引文数据,以 6 年为计量时间窗口,依析取的 8 个外部特征因素计算不同分组的未被引率,采用非参数方法检验各因素是否存在显著差异。[**结果/结论**]8 个外部特征因素对论文未被引都有显著影响,其中作者所属机构的影响相对较小,作者发文时年龄与论文篇幅的影响相对较大,作者发文时职称、作者数量、参考文献数量、关键词数量、基金类别的影响程度大致相仿;各因素的未被引率在前 3 年的变化较为剧烈,后 3 年变化较为平缓;各因素未被引率的时间序列变化趋势各不相同,其影响平稳性也变化各异。

关键词:未被引论文 零被引论文 图书馆情报与文献学 文献外部特征 作者因素 论文因素 基金因素

分类号:G250.252 G301

DOI:10.13266/j.issn.0252-3116.2018.04.001

1 引言

20 世纪 50 年代 E. Garfield 创见性提出引文分析方法,引文分布分析(citation distribution analysis)就广泛应用于科研绩效、期刊和个人影响等的定位、比较和评估,是计量学的核心研究主题[1]。目前其研究主要聚焦于引用分布曲线上代表"高被引和受关注"的头部,极少关注"低被引和暂时无人关注"的尾部。然而,"任何学科或任何地方出版的论文中,总会有一些从未受到任何引用的论文"是科学界中普遍存在且不可忽视的事实[2]。

长尾理论认为,代表总量 80% 的尾部可能产生与 20% 头部相当的效益。所以,在当前科研环境数字化与网络化语境下,代表文献长尾部分的低被引或未被引文献对科研的贡献可能与头部高被引文献的贡献相当,或至少对推动科学研究具有重要作用;而且,在一定时间窗口中未被引状态的文献并不一定全是低质量或没有学术价值,其中也不乏潜在"精品",如一些"迟滞承认"和"睡美人"文献[3]。故而对论文未被引现象的研究是引文分布研究中不可或缺的部分。

因此,重点关注"头部"而忽略"尾部"的研究实践不能反映科学交流的全貌,未被引研究对情报科学具有重要意义。一方面,只有透彻研究普遍意义的未被引文献规律,才能完整认知引文分布谱系,因此论文未被引现象研究有利于丰富和扩展计量学的研究范畴;另一方面,通过对未被引影响因素的研究,探寻科研人员的引用选择规律,对于识别未被引文献的产生机制和最大限度避免科研资源浪费与提升科学交流效率具有重要意义。

2 相关研究工作

很早有学者就注意到论文未被引的原因多种多样,不仅仅是质量原因,还存在许多因素造成论文没有受到关注和没有被引[4]。未被引现象得到大规模研究出现在 20 世纪 90 年代,E. Garfield 指出论文未被引的影响因素很多,可能是由于论文太平凡、质量不高,或者主题较偏,甚至可能是论文尚未被吸引进当前的知识范式、研究内容过于新颖超前或者遭遇迟滞承认等

作者简介:韩毅(ORCID:0000-0001-7021-3229),教授,博士,博士生导师,E-mail:hanyi72@swu.edu.cn;伍玉,硕士研究生;申东阳,硕士研究生;况书梅,硕士研究生;袁庆,硕士研究生。

收稿日期:2017-08-31 修回日期:2017-11-22 本文起止页码:6-13 本文责任编辑:王善军

因素造成的[5]。

大部分相关研究工作聚焦于论文本身因素来探索论文的未被引规律。R. E. Stern 通过比较被引文献和未被引文献的作者数量、关键词数量、标题词数量、参考文献数量、期刊的年龄和价格等特征值后发现,参考文献数量对论文是否被引的影响最大,而其他特征对论文未被引的影响较小[6]。R. Rousseau 基于观测数据发现:科学交流系统中有很多文献与某作者的研究工作相关,但作者仅会在自己的研究成果中将极少的一部分列为参考文献;通过贝叶斯概率模型测算出仅有一个作者的论文被引概率仅为 8%,而多作者的合作论文被引概率则高达 33%[7]。胡泽文等以 6 种高影响力期刊 1992-1999 年发文为研究对象,统计这些期刊不同长度的论文数量及不同长度论文的未被引率,结果发现:短论文未被引率非常高,长度在 1-4 页的文章未被引占比非常大,相较而言长论文获得引用的可能性更大[8]。石磊引入生存分析理论讨论基金项目、作者规模、论文篇幅和参考文献数量对未被引论文生存状况的影响,发现是否获基金资助、论文篇幅对未被引论文生存有显著影响[9]。温芳芳通过未被引与高被引论文的比较发现作者影响力与论文未被引成反向相关,基于未被引与高被引论文关键词共现网络发现两者的核心研究点存在较大差异[10]。

以期刊为载体的未被引研究也是学者们关注的重点。实证研究发现:期刊影响因素作为未被引的重要影响因素,其影响表现是 S 曲线,其拐点是影响因子平均值和未被引因子平均值[11];刘娜以军队医学期刊为研究对象,详细分析后发现:期刊主办单位、期刊影响力、期刊影响因子等特征对期刊论文未被引有重要影响[12]。

探索不同学科、不同国家的未被引论文分布差异是更高层面的关注焦点。以中国科技核心期刊为对象探索未被引论文的学科分布和机构分布发现:社会科学未被引论文比例大大高于自然科学,高校机构发文未被引率低于平均水平,研究院发文未被引率在所有各类型机构中最低[13]。对比分析中国和印度国际合作论文和非国际合作论文的未被引率发现:国际合作关系结构是影响未被引率的重要因素,合作国家越多未被引率越低[14];印度论文的篇均被引频次是影响论文未被引率的主要因素,中国论文的未被引率受篇均参考文献数量影响较大,其次是篇均作者数量,而篇均被引频次的影响因学科而异[15]。

根据相关研究成果,可归纳出论文未被引的影响因素集(见表 1)。可以看到:现有关于未被引现象的主要研究较为全面地反映了未被引的基本特征,但各项具体研究工作仅就可能的几个有限因素影响进行了分析,没有对所有可能因素进行系统探讨,且研究结论还存在诸多不一致;研究方法主要以固定时间窗口应用数理统计、相关性计算和散点图拟合等来分析未被引特征,没有从时间维度来探索可能的演化趋势;所有未被引影响因素主要集中于文献外部特征,较少从内容特征来分析未被引的基本属性。

表 1　论文未被引的影响因素集

影响因素类别	影响因素项目
作者影响因素	作者数量[7,14-15],作者所属机构[13-14]
论文影响因素	关键词数量[10]、参考文献数量[6,14]、论文篇幅[8-9]
其他影响因素	基金类别[9],期刊类别[11-12],学科[13],国别[14]

本文将以中文科研论文为对象,综合已有研究成果析取未被引的外部特征,从时间序列揭示各因素的影响趋势,识别出中文科研论文未被引的主要影响因素,进而为探讨未被引现象产生的机理机制奠定基础。

3　数据与方法

3.1　数据获取

计量学研究结果强烈依赖于所获取的数据,为了使研究结论具有可比性,应尽可能选择大家公认、数据全面的权威信息源,本文选择 CSSCI 为来源数据库;为了便于研究结果的解读,本文选择图书馆情报与文献学作为样本学科。

基于上述选择结果,排除表 1 中的“国别”“学科”影响因素;同时,CSSCI 收录的都是各学科重要核心期刊,也不便于对之进行期刊类别划分,故排除“期刊类别”因素。根据 CSSCI 提供的数据资料及一些论文具体形态,根据经验判断与作者相关的两个因素——发文时职称和发文时年龄,可能会对论文未被引有影响,故把这两个因素也添加到影响因素集中。根据小规模数据探索结果,最终确定中文未被引论文的外部特征影响因素集及相应的指标取值(见表 2)。

本文数据采集于 2016 年 11 月。分析时期望以整年数据为对象进行对比,由于 2016 年 CSSCI 收录引文数据不全,故在分析论文未被引外部特征因素影响时仅以 1998-2015 年发表的期刊论文为样本数据。

具体的数据收集策略如下:

(1)综合图书馆情报与文献学学者们的 h 指数特征[16],以随机方式选择 200 名学者为样本对象;

(2)从 CSSCI 获取以样本学者为第一作者于 1998-

表 2 影响中文科研论文未被引的外部特征因素及指标取值

影响因素类别	主要指标	指标取值细分
作者影响因素	作者发文时职称	正高级、副高级、其他
	作者发文时年龄	36 岁及其以下、37 – 42 岁、43 – 48 岁、49 – 54 岁、55 – 60 岁、60 – 66 岁、67 岁及其以上
	作者所属机构	科研院所、高校教学单位、高校其他单位、其他单位
	作者数量	1 个、2 个、3 个、4 个及以上
论文影响因素	关键词数量	1 – 3 个、4 个、5 个、6 个及以上
	参考文献数量	1 – 5 条、6 – 10 条、11 – 15 条、16 条及以上
	论文篇幅	1 – 3 页、4 – 5 页、6 – 7 页、8 页及以上
其他影响因素	基金类别	1 类、2 类、3 类、4 类基金论文，非基金论文

2015 年期间发表期刊论文的详细信息，以 txt 文件方式存储下载的 CSSCI 各数据项；

(3) 由于一些外部特征因素数据，如作者发文时职称、年龄等，无法从数据库中直接获取，因此通过公开途径获得样本学者的出生日期和晋职时间；

(4) 自编程序从 txt 文件中提取各字段数据，以及作者发文时职称、年龄数据，分别以 excel 文件存储。同时，分别获得一个包括所有样本论文题名的 txt 文件与一个所有样本论文作者名的 txt 文件，两个文件中论文题名与论文作者依次对应，将两个 txt 文件导入自编爬虫程序中，根据题名和第一作者获取各篇论文的被引数据，并以 excel 文件方式存储。

3.2 研究方法

为了识别出表 2 中各外部特征因素的影响效果，本文通过比较各因素不同类别的未被引率是否存在显著差异。

所谓未被引率 (Rate of Uncited Articles, RUCA)，是指样本论文集在出版后的某个特定时间窗口中从未被引用的论文数量占该论文集总数量的比例。考虑到图书馆情报与文献学领域引用半衰期大约为 4 年[16]，为了充分消除引用延后的影响，本文选取 6 年为被引时间窗口，以在足够长的时间窗口内观察未被引率的变化趋势。需要说明的是，时间计量单位是以自然年度为标准，论文发表当年即为第 1 引用年。

未被引率的计算公式如下：

$$RUCA = (1 - C_{nj}/N_i) * 100\%$$

其中，i = 1998, 1999, …, 2010, j = 1, 2, …, 6, N_i 为第 i 年发表的文章总数，C_{nj} 为论文在发表后第 j 年已经被引用过的文章数量。

具体的数据处理与分析步骤如下：

(1) 对应整理样本论文集的各项信息及引用数据，以 excel 文件方式存储；

(2) 计算样本论文集发文后 1 – 6 年各外部特征因素不同细分类别的未被引率，绘制未被引率的时间演化图，由于数据少不符合 t 检验或方差分析要求，故采用配对样本非参数检验方法检验不同细分类别是否存在显著差异。

4 研究结果

4.1 作者因素对论文未被引的影响

4.1.1 作者职称对论文未被引的影响 在数据分析时发现中级职称、初级职称以及在校学生发文量相对较小，并且小样本数据分析发现这些作者发文未被引率差异并不十分显著，故将中级职称、初级职称与在校学生归为一类。正高级职称包括教授及相应职别，副高级职称包括副教授及相应职别。将样本数据以职称分三组分别计算发文后 1 – 6 年的未被引率并绘制其演化图，使用 SPSS 进行配对样本非参数检验。

从图 1 可以发现：正高级职称与副高级职称发文未被引情况略微存在差异，而高级职称与其他类职称的差异较为明显，整体上呈现出职称越高其发文未被引率越低。发文后第 1 年，不同职称作者发文未被引率相近；发文第 2 年及以后，职称越高其未被引率越低，正高级职称未被引率略低于副高级职称的未被引率；第 4 年开始副高级职称与正高级职称未被引率演化趋势趋同。根据对样本科研人员的分析，大部分科研人员在副高级职称过后 5 年左右晋升为正高级职称，即可能是发文 5 年后职称的变化使得正高级职称作者发文与副高级职称作者发文未被引率演化趋势趋同。

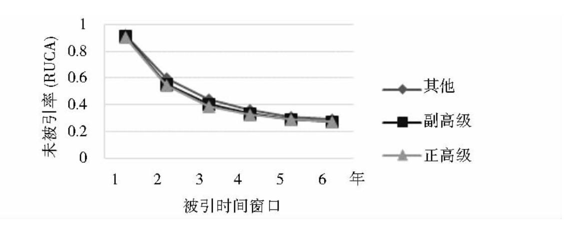


图 1 不同职称作者发文后 1 – 6 年的未引率演化

从表 3 可以发现：卡方统计量为 12.000，W 系数为 1.000，相伴概率为 0.002，小于显著水平 0.05，可认为三组数据存在显著差异，即作者职称对论文未被引有显著影响。

表3　不同作者因素发文后1－6年的未被引率差异检验

	发文职称	发文年龄	所属机构	作者数量
N	6	6	6	6
Kendall W[a]	1.000	.760	.478	.767
卡方	12.000	27.357	8.600	13.800
df	2	6	3	3
渐近显著性	.002	.000	.035	.003

注：a 为 Kendall 协同系数

4.1.2　作者年龄对论文未被引的影响　采用相同方法得到作者年龄对未被引的影响结果。从图2可以发现：各年龄段作者发文在1－6年时间窗口其未被引率呈现逐渐下降趋势。其中，36岁及以下、37－42岁与67岁及以上年龄段作者发文在各被引时间窗口中相较于其他年龄段的未被引率始终处于较高水平，尤其是67岁及以上年龄段始终处于高未被引率状态；43－66岁年龄段的几个类别则呈现较低未被引率，在第2年和第3年被引时间窗口时尤为突出。更为细致的分析发现：在发文当年，各年龄段发文未被引率差异并不明显；发文后第2年和第3年，43－48岁年龄段与55－60岁年龄段出现未被引率低谷，即这两个年龄段发文整体呈现出较好的引用态势；发文后第4年及以后，43－48岁年龄段与36－42岁年龄段以及67岁及以上年龄段的论文未被引率差异不再明显，而55－60岁、61－66岁年龄段的未被引率明显低于其他年龄段，其中61－66岁年龄段的未被引率较前3年被引时间窗口有显著下降，形成了新的未被引率低谷。

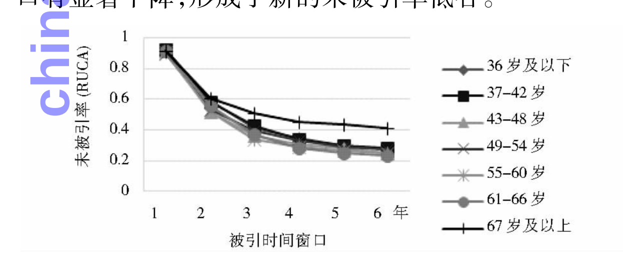


图2　不同年龄作者发文后1－6年的未被引率演化

表3非参数配对检验结果显示：卡方统计量为27.357，W系数为0.760，相伴概率为0.000，小于显著水平0.05，可认为各年龄段的发文未被引率存在显著差异，即论文未被引受到作者发文年龄的显著影响。

4.1.3　作者所属机构对论文未被引的影响　作者所属机构类型划分为4种：科研院所，包括各类信息研究所、情报研究所等；高校教学单位，即高校各院系；高校其他单位，包括各高校图书馆、档案馆、研究中心等；其他单位，包括杂志社、编辑部、公司企业等。采用相同的方法得到作者所属机构类型对未被引的影响结果。

从图3可以看到：4类机构发文未被引情况存在较大差异。从发文后第2年开始科研院所发文未被引率一直处于4类机构中最低水平，发文后前3年未被引率快速下降；高校教学单位和高校其他单位发文未被引率前2年有差异，从第3年开始趋同；其他类的未被引率一直处于最高水平，且与其他几类机构发文未被引率差异较大，在发文后引用高峰期的2－3年其未被引率下降幅度也不大。

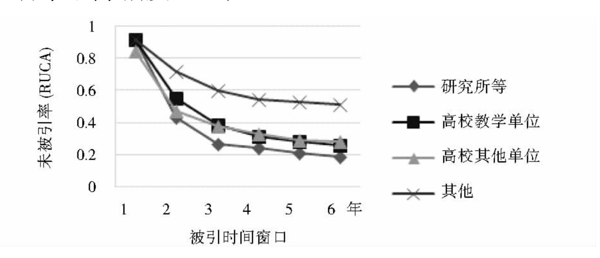


图3　不同机构作者发文后1－6年的未被引率演化

表3非参数配对检验结果可看到：卡方统计量为8.600，W系数为0.478，相伴概率为0.035，小于显著水平0.05，可认为各机构发文未被引率存在显著差异，即作者所属机构属性对论文引用与否有显著影响。

4.1.4　作者数量对论文未被引的影响　研究表明，科研合作可显著提高论文的引用率[17]，通过合作不仅可以实现科研人员之间的分工以提高效率，也有利于通过科研人员之间的知识交流提高论文质量。但合作论文与独著论文在引用上是否存在较大差异，以及是否作者规模越大论文被引情况越好，需要有数据的进一步证明来确认。采用相同的方法得到论文作者数量对未被引的影响结果。

从图4可以发现：整体上，不同作者数量的科研论文零被引率均逐年递减，但不同作者数量的论文未被引率差异在发文前两年并不明显，而从第3年开始作者数量越多未被引率越低。其中，1个作者论文1－6年的未被引率始终处于较高水平，尽管在发文后第1年至第2年与其他类别的未被引率差异并不明显，但从第2年开始随着时间推移其差异逐渐增大；2个作者与3个作者论文的未被引率在1－4年被引时间窗口内差异并不明显；4个及以上作者科研论文未被引率与1－3个作者的未被引率在1－4年被引时间窗口内存在差异，但从发文后第4年开始与2个和3个作者的未被引差异逐渐减小。

表3的数据显示：卡方统计量为13.800，W系数为0.767，相伴概率为0.003，小于显著水平0.05，可认为不同作者数量论文的未被引率存在显著差异，即论

文未被引受到论文作者数量的显著影响。

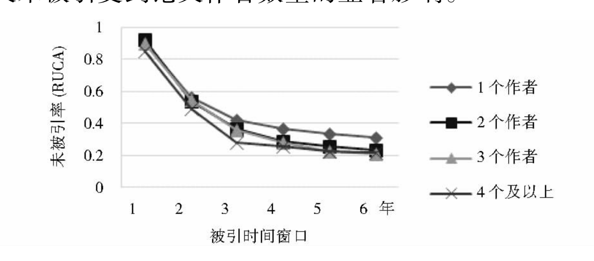


图 4　不同作者数量发文后 1 –6 年的未被引率演化

4.2　论文因素对论文未被引的影响

4.2.1　关键词数量对论文未被引的影响　关键词是作者对论文研究内容的高度概括，关键词多寡直接影响检索系统中论文的检出几率。本文按照关键词数量分为 4 类：1 –3 个关键词、4 个关键词、5 个关键词、6 个及以上关键词。采用相同的方法得到关键词数量对论文未被引的影响结果。

从图 5 可以发现：总体上讲，论文关键词越多其未被引率越低，但关键词数量不同组别的论文未被引率差距并不十分明显，只有 1 –3 个关键词类别的论文未被引率与 6 个及以上关键词的有一定差异，但在发文第 3 年及以后，不同数量关键词类别的论文各年未被引率并无明显差异，即引文峰值年前关键词数量才有显著性意义。

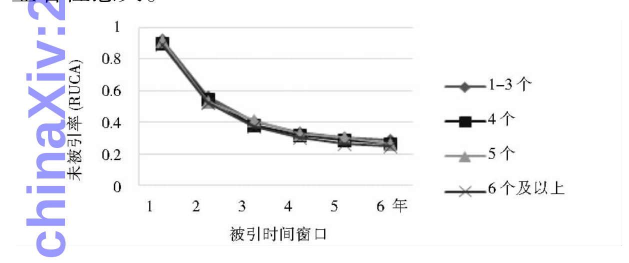


图 5　不同数量关键词类别发文后 1 –6 年的未被引率演化

表 4 检验数据显示：卡方统计量为 15.000，W 系数为 0.833，相伴概率为 0.002，小于显著水平 0.05，认为关键词数量对论文的未被引率有显著差异，即论文引用与否受到关键词数量多少的显著影响。

表 4　不同论文因素与基金类别发文后 1 –6 年的未被引率差异检验

	关键词数量	参考文献数量	论文篇幅	基金类别
N	6	6	6	6
Kendall W[a]	.833	.722	1.000	.811
卡方	15.000	17.333	18.000	19.467
df	3	4	3	4
渐近显著性	.002	.002	.000	.001

注：a 为 Kendall 协同系数

4.2.2　参考文献数量对论文未被引的影响　参考文献是科研工作的联系性、继承性与相关性体现，也是作者对当前研究内容在已有知识体系中位置的认知反映，一定程度可反映论文的质量与信息量。本文按照参考文献数量将论文分为 4 类：1 –5 篇参考文献组、6 –10 篇参考文献组、11 –15 篇参考文献组、16 篇及以上数量参考文献组。采用相同的方法得到不同参考文献数量对未被引的影响结果。

从图 6 中可以看出：不同参考文献数量论文的未被引率差异较为明显且变化趋势较为一致，参考文献数量越多的论文其未被引率越低。发文第 1 年不同参考文献数量论文的未被引率差异较为明显；第 1 年至第 2 年，参考文献越多论文的未被引率变化越大。

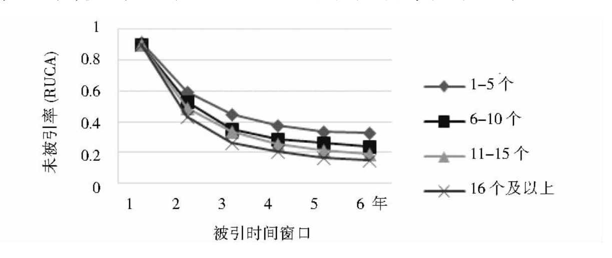


图 6　不同参考文献数量论文发文 1 –6 年的未被引率演化

表 4 的非参数检验结果表明：卡方统计量为 17.333，W 系数为 0.722，相伴概率为 0.002，小于显著水平0.05，可认为不同参考文献数论文的未被引率存在显著差异，即论文引用与否受到参考文献数量的显著影响。

4.2.3　论文篇幅对论文未被引的影响　由于难以直接获取论文字数，且一般论文多含图表等内容，因此本文选取论文页数作为论文长度的度量指标，以此探究论文篇幅对论文未被引的影响。根据多次探索实验，将论文篇幅划分为 4 类：1 –3 页组、4 –5 页组、6 –7 页组、8 页及以上组。采用相同的方法得到不同论文篇幅对未被引的影响结果。

图 7 显示：整体上，不同被引时间窗口中论文篇幅越大未被引率越低；随着论文篇幅增加，被引时间窗口中论文未被引率逐渐降低，且篇幅越长论文在发文后第 2、3 年未被引率下降越快。其中，篇幅 1 –3 页论文的未被引率一直保持在较高水平，与其他类别论文的差异明显；篇幅 4 –5 页论文的未被引率保持中等水平；篇幅 6 –7 页论文与 8 页及以上论文的未被引率差距并不十分明显，但与其他类别的未被引率存在较大差异。从时间动态性来看，发文后第 1 年不同篇幅论

文的未被引率已呈现差异,论文篇幅越长未被引率越低;发文后第2年,论文篇幅越长未被引率下降幅度越大,随着时间推移其降幅减弱。

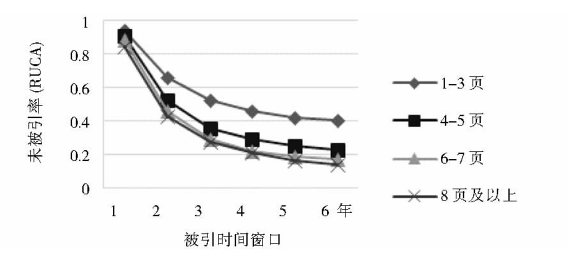


图7　不同篇幅论文发文1－6年的未被引率演化

表4非参数检验结果表明:卡方统计量为18.000,W系数为1.000,相伴概率为0.000,小于显著水平0.05,可认为不同论文篇幅论文的未被引率存在显著差异,即篇幅长短对论文是否引用有显著影响。

4.3　基金类别对论文未被引的影响

本文将基金划分为4类:第1类包括国家社科基金重大重点项目、国家软科学研究计划重大项目、国家自然科学基金重大重点项目,第2类包括国家社科基金项目、国家软科学研究计划项目、国家自然科学基金项目,第3类包括国家部委、各省(自治区、直辖市)基金项目,第4类包括其他各类基金项目。相应的论文划分为5个类别,其中第5类为非基金项目论文。采用相同的方法得到不同基金类别对论文未被引的影响结果。

图8表明:不同类别基金论文在1－6年被引时间窗口的未被引率逐渐下降,但同比来看并不完全是基金类别越高未被引率越低。第1类基金论文未被引率在发文后前3年居于最低水平,第3类基金论文未被引率在发文后第5年和第6年处于最低水平;第1类基金论文与第2类基金论文未被引率在第4年和第5年逐渐趋同;第4类基金论文和非基金论文未被引率始终保持在较高水平。

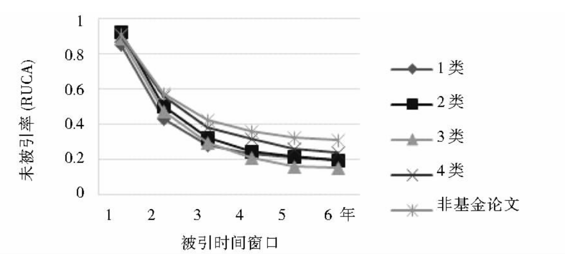


图8　不同基金类别论文发文后1－6年的未被引率演化

表4配对样本非参数检验结果显示:卡方统计量为19.467,相伴概率为0.001,小于显著水平0.05,可认为不同组别论文的未被引率存在显著差异,即论文是否得到基金资助及类别不同会显著影响论文的引用状态。

5　结论与讨论

5.1　主要结论

分析结果显示,本文所讨论的三大类因素8个指标对中文论文未被引均有显著影响。从检验的相伴概率值来看,作者所属机构对论文未被引的影响相对较小,作者发文时年龄与论文篇幅对论文未被引的影响相对较大,其他5个因素的影响程度大致相仿。

总体上看,各类影响因素的未被引率在前3年的变化较为剧烈,后3年变化较为平缓,表明图情学科论文的引用高峰通常在出版后的3年左右,不同影响因素的具体程度略有不同。

从未被引率的时间序列变化来看,有的影响因素在整个时序上其差异越来越小,有的影响因素其差异越来越大,但更多因素表现出较为稳定的差异特性;而且,不同影响因素在同一时点上的未被引率差异变化较大,其差异值有的表现得较为平稳,有的表现得极不稳定,更多表现出是部分平稳特征。具体结果如表5所示:

表5　论文外部特征因素对未被引的影响差异

影响因素	差异显著性(相伴概率值)	时间序列趋势性	因素影响平稳性
作者发文时职称	显著(0.002)	差异减小	平稳
作者发文时年龄	显著(0.000)	差异稳定	非平稳
作者机构	显著(0.035)	差异增大	部分平稳
作者数量	显著(0.003)	差异增大	部分平稳
关键词数量	显著(0.002)	差异稳定	部分平稳
参考文献数量	显著(0.002)	差异稳定	平稳
论文篇幅	显著(0.000)	差异稳定	部分平稳
基金类别	显著(0.001)	差异稳定	非平稳

5.2　讨论

本文以CSSCI数据库提供的数据为基准,根据析取的三大类因素8个指标对论文未被引率的影响进行了探索,发现8个因素对未被引率都有显著影响,其中作者机构[13]、作者数量[7]、关键词数量[6]、参考文献数量[6]、论文篇幅[8－9]、基金类别[9]等的影响再次得到证实,本文提取的作者发文时年龄及职称也对未被引有显著影响,这与相关研究中提及的作者影响力有直接关联[10]。无论对期刊、作者还是其他相关利益者,这些显著性影响因素对于期刊选稿、论文作者排序、论文

关键词标注、参考文献选择等都具有指导性作用。

相较于已有研究而言,本文的主要贡献是揭示了各影响因素不同组分对论文未被引影响的时间维度变化特征。过去的研究更多是在时间节点上发现这些因素的显著影响作用,本文展示了 6 年范围内各因素组分的时间演化特征。这些特征显示了随着时间的推移,每个影响因素不同分组数据的未被引率在不同时点上的绝对值及差异,以及这些差异在时间序列上的演化趋势特征,其中差异增大的因素包括作者机构与作者数量,差异减小的因素是作者发文时职称,其他因素的差异较为稳定。这些差异性特征为显著性影响的最优组分识别提供了可视化结果,在一定程度上可解决非参检验只有检验结果而没有细节性信息展示的缺陷。

影响论文未被引的因素非常多,囿于本文的数据来源和单一样本对象,本文结论仅是关于中文科研论文未被引的影响因素识别,结论普适性有待进一步验证;由于本文时间序列数据的有限性,本文仅采用非参数配对检验来比较不同组分的差异性,未能把这些不同影响因素综合起来并考虑彼此间相互作用情形下的全面分析。在未来研究中,需要选择更多的样本学科、不同语种、不同国别、不同来源期刊的更多数据来探索更完整的未被引影响因素集,应用面板数据分析方法确定各影响因素的重要程度及每个因素的显著水平。

参考文献:

[1] STRINGER M J, PARDO M S, AMARAL L A N. Statistical validation of a global model for the distribution of the ultimate number of citation accrued by papers published in a scientific journal[J]. Journal of the American Society for Information Science, 2010, 61(7):1377 - 1385.

[2] BURRELL Q L. Will this paper ever be cited? [J]. Journal of the American Society for Information Science and Technology, 2002, 53(3):232 - 235.

[3] 李江. 科学中的"睡美人"与"昙花一现"现象评述[J]. 大学图书馆学报, 2016, 34(3):38 - 43.

[4] GHOSH J S. Uncitedness of articles in nature, a multidisciplinary scientific journal [J]. Information processing & management, 1975, 11(5):165 - 169.

[5] GARFIELD E. To be an uncited scientist is no cause for shame [J]. The scientist, 1991, 5(6):12.

[6] STERN R E. Uncitedness in the biomedical literature[J]. Journal of the American Society for Information Science, 1990, 41(3):193 - 196.

[7] ROUSSEAU R. Why am I not cited, or why are multi-authored papers more cited than others? [J]. Journal of documentation, 1992, 48(1):79 - 80.

[8] HU Z, WU Y. Regularity in the time-dependent distribution of the percentage of never-cited papers: an empirical pilot study based on the six journals[J]. Journal of informetrics, 2014, 8(1):136 - 146.

[9] 石磊. 期刊论文零被引现象实证研究[D]. 蚌埠:安徽财经大学, 2016.

[10] 温芳芳. 我国情报学期刊论文零被引的成因及影响因素探析[J]. 情报理论与实践, 2016, 39(4):27 - 31, 26.

[11] EGGHE L. The distribution of the uncitedness factor and its functional relation with the impact factor[J]. Scientometrics, 2010, 83(3):689 - 695.

[12] 刘娜. 军队医学期刊的零被引现状及其影响因素研究[D]. 北京:中国科学技术信息研究所, 2015.

[13] 朱梦皎. 科技论文零被引现象探析与实证[D]. 北京:中国科学技术信息研究所, 2013.

[14] 郭永正. 国际合作论文零被引率的中印比较[J]. 情报杂志, 2014, 33(12):89 - 93.

[15] 郭永正. 非国际合作论文零被引率的中印比较[J]. 图书与情报, 2015, 34(4):90 - 95.

[16] 韩毅, 夏慧. 时间因素视角下科研人员评价的 Pt 指数研究[J]. 中国图书馆学报, 2015, 41(6):73 - 85.

[17] FIGG W D, DUNN L, LIEWEHR D J, et al. Scientific collaboration results in higher citation rates of published articles[J]. Pharmacotherapy, 2006, 26(6):759 - 767.

作者贡献说明:

韩毅:整体研究设计与规划,论文修改完善;
伍玉:数据搜集、整理与分析,撰写论文初稿;
申东阳:程序编写,数据搜集;
况书梅:参与数据搜集、整理与分析;
袁庆:参与数据搜集、整理与分析。

Part Ⅰ of the Exploration on Uncited Papers in Chinese: The Influences of External Features of Papers - A Case Study of Library and Information Science

Han Yi　Wu Yu　Sheng Dongyang　Kuang Shumei　Yuan Qing
College of Computer and Information Science, Southwest University, Chongqing 400715

Abstract: [**Purpose/significance**] It is of importance to explore the uncited phenomenon, which is essential to the citation distribution. It not only enriches and expands the metrics fields, but helps to reveal the generative mechanism for

uncited papers, making the maximum use of research resources, and enhancing the efficiency of the scientific communication. [**Method/process**] CSSCI was chosen as the source database. Library and information science was selected as the sample, 200 scholars were chosen randomly according to the distribution features of h index, and their papers and responding citation data were retrieved from the database. This paper took 6 years as the time window. 8 external features were extracted, and the rate of uncited papers was calculated for every group. Non-parametric tests were used to examine factorshaving significant differences. [**Result/conclusion**] 8 external features have significant impact on uncited papers, in which the influence of authors' affiliation is relatively smaller, the influence of authors' age when they published papers and the length of papers are relatively larger. The influence of other 5 factors, namely, authors' title, author numbers, reference numbers, keywords numbers, and fund categories, is similar to each other. In the first 3 years, the changes of the rate of uncited papers are much sharper, and in the second 3 years, the changes are relatively gentle. The time series trend and the influence stability of the rate of uncited papersvary.

Keywords: uncited articles　non-cited articles　library and information science　external features of papers　author-related factors　paper-related factors　fund-related factors